Reviewer Comments:   
Reviewer #1:   
  
Suitable Quality?: Yes   
Sufficient General Interest?: Yes   
Conclusions Justified?: Yes   
Clearly Written?: Yes   
Procedures Described?: No   
Supplemental Material Warranted?: Yes   
  
Comments:   
This is a really excellent and thorough revision to an already exciting paper and I congratulate the authors. The new experiments significantly extend the interpretation of the result, ruling out alternatives I actually didn't think were possible to fully address. Just to reiterate, this is an incredibly careful and targeted behavioral attack on the mechanisms of approximate planning in the brain, which is an issue of strong current interest, and goes well beyond previous attempts in this area in its specificity.   
  
I have only a few suggestions, only one of any import:   
  
(1) It seems like some methodological information might have gotten lost in the reorganization (or I have lost track of it anyway?). I don't see a discrete methods section anywhere and I can't find things like the numbers of subjects and trials, the fact that the subjects were recruited from Turk, an assertion that informed consent was obtained, and the process by which random rewards were generated. I think I have reviewed all that information previously and much of it also appears in the response letter, so it's not an issue for me but of course it should appear in any published version.   
  
Minor:   
  
(1) I think it's worth briefly comparing to the results of Dezfouli & Balleine (PLoS CB 2013). Does the current model explain their action sequencing effect? I think one key difference is that they are lacking an experiment like Expt 2 here and so can't really substantiate their claim that option selection is model-based.   
  
(2) The control for win-stay-lose-shift seems a bit fishy to me since it doesn't appear to control for the reward obtained on the first trial after the setup (which is presumably correlated with the regressor of interest). Personally I don't find this control at all necessary in any case since win-stay-lose-shift is to my mind a bona fide instance of model-free learning (it just arises in the limit as learning rates approach 1).   
  
  
  
Reviewer #2:   
  
Suitable Quality?: Yes   
Sufficient General Interest?: No   
Conclusions Justified?: Yes   
Clearly Written?: Yes   
Procedures Described?: Yes   
Supplemental Material Warranted?: Yes   
  
Comments:   
The authors have taken care to address many of the issues raised in the last review. However, there are several sticking points for this reviewer.   
  
1. On the assertion that model fitting as requested in too difficult. There are several papers from Bernard Balleine that also argue for hierarchy in these two-steps tasks, but they assert the opposite: the highest level of option selection is controlled by a model-based system, and the internal option policies are model-free.   
  
Cushman cites one of these papers in passing (ref #32), but doesn't discuss this connection. This is important. Also, Balleine's group does formal model fitting with option models, so the assertion in the first reply that model fitting here is too hard or the data too noisy is hard to understand.   
  
2. What I am mostly puzzled about though is that it seems like they set up a straw man type argument, essentially asking: "is there evidence of model-free behavior at the top level?" (and secondary to that, whether the problem representation is hierarchical), whereas what they want to ask is "is there ONLY model-free behavior at the top level, and no model-based control at all?" This seems pretty silly, so maybe I missed something in my re-reading. In exp 1a/b, the fact that rewarded rare transitions are reinforced is taken as evidence for Model-Free. Yes. But it's not evidence against any possible Model-based influence.   
  
3. Then there is the secondary question of whether the problem representation is hierarchical and exactly where the option boundaries are, which exp 2 is meant to address. The focus is on what constitutes a "goal state". However, the other extremely important piece of what constitutes an option -- the the option policy -- is completely trivial. It's not even a two-step problem, it's one step: 1,3 maps to blue state, 2,4 to red state. The experiment, as designed, can't speak strongly as to whether or not the same option policies are being invoked in each instance. I point these out because this was not highlighted by the authors in either version of the ms.   
  
(4,5 are minor)   
  
4. In the SI, why do t-tests on model output which can be made arbitrarily precise by running a larger number of simulations? This SI is more akin to a power analysis than actually testing the predictions of the models. Unless I'm missing something here, the results speak against them because even without model-free goal learning the difference between the quantities of interest are close to "significance" at 0.1.   
  
5. SI equation 4 is a bit confusing. I understand they want to unroll the tree, but what is "a"? In the equation it says it is the action set which leads to a goal state (so actions at the first stage, since goal states are at the second stage?), but above it says "from each stage 2 action a".